



BANA 4090: Chapter 4: ARIMA models (Part IV)

Zhaohu(Jonathan) Fan

07/15/2021

- Main topics:
 - Seasonal ARIMA models

Seasonal ARIMA models

ARIMA models

- The general **idea** of ARIMA models is to **capture autocorrelation**.
 - **Autocorrelation**: measure linear relationship between **lagged values** of a time series y .
 - For example, we measure the relationship between:
 - y_t and y_{t-1}
 - y_t and y_{t-2}
 - etc.
- **Major assumption**: **stationarity**.
- **Advantages**: strong underlying theory, flexible, etc.
- **Key concepts**: **order, differencing**.
 - Autoregressive Integrated Moving Average models (ARIMA (**p, d, q**) models):
 - AR: p = **order of the autoregressive part**.
 - I: d = **degree of first differencing involved**.
 - MA: q = **order of the moving average part**.

Stationarity

- First-order (Lag-1) differencing (used for removing trend)
 - The differenced series is the **change** between each observation in the original series:
 - $y'_t = y_t - y_{t-1}$. In R, `diff(data, lag=1)` or `diff(data)`.
- Seasonal (Lag-m) differencing (used for removing seasonality)
 - is the difference between an observation and the corresponding observation from the previous year.
 - $y'_t = y_t - y_{t-m}$, where m= number of seasons.
 - For monthly data m=12. In R, `diff(data, lag=12)`.
 - For quarterly data m=4. In R, `diff(data, lag=4)`.

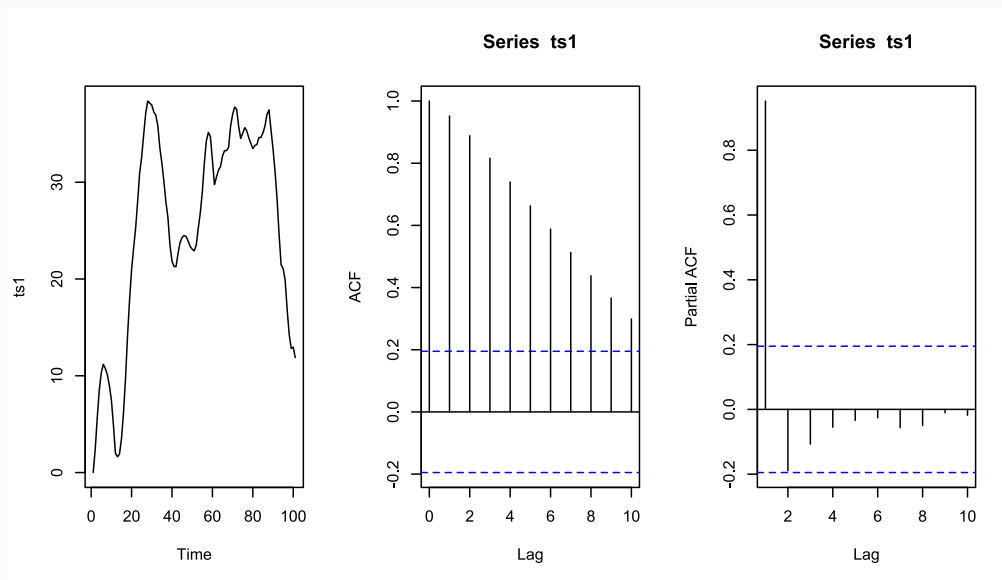
Stationarity (cont'd)

- Augmented Dickey-Fuller Test (ADF)
 - In general, the Augmented Dickey-Fuller test is defined as:
 - H_0 (null hypothesis): the data series is not stationary
 - H_a (alternative hypothesis): the data series is stationary.
- The `adf.test()` from the R package `tseries` will do a Augmented Dickey-Fuller test.

Identifying ARIMA (p, d, q) models

Model	ACF	PACF
AR(p)	exponentially decay	cut off at lag p
MA(q)	cut off at lag q	exponentially decay
ARMA(p,q)	exponentially decay after lag q	exponentially decay
ARIMA(p,d,q)	slowly decrease	exponentially decay

- ARIMA (1,1,1)
- Note: **ACF** starts at lag 0 and **PACF** starts at lag 1.



Seasonal ARIMA models

- (Non-seasonal) ARIMA models: non-seasonal data.
- ARIMA models are also capable of modelling a wide range of **seasonal data** by including additional **seasonal terms**.
- Seasonal ARIMA models:

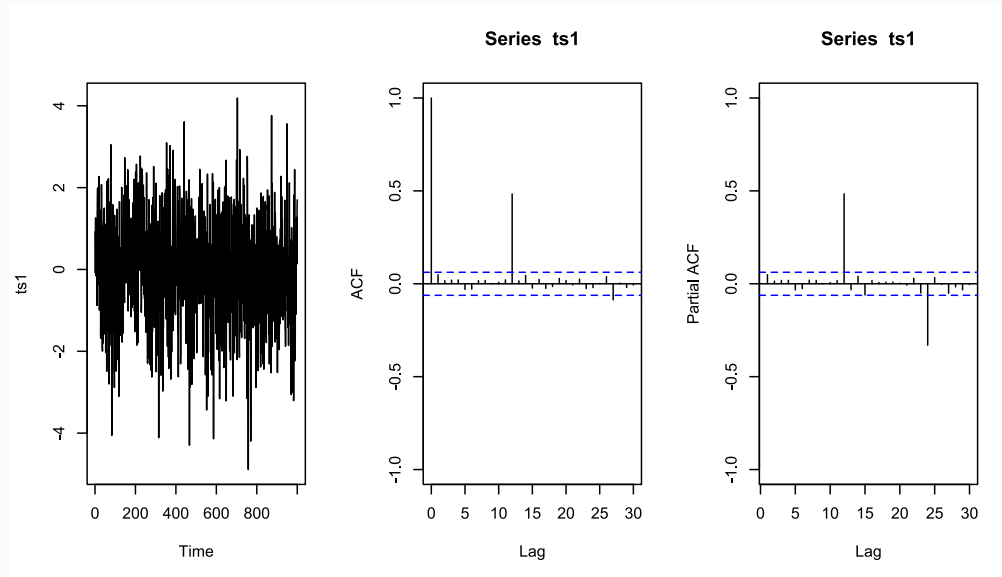
ARIMA	(p, d, q)	$(P, D, Q)_m$
	$\underbrace{\hspace{1cm}}$	$\underbrace{\hspace{1cm}}$
	↑	↑
	Non-seasonal part	Seasonal part of
	of the model	of the model

where m = number of seasons/number of observations per year.

- For monthly data $m=12$.
- For quarterly data $m=4$.

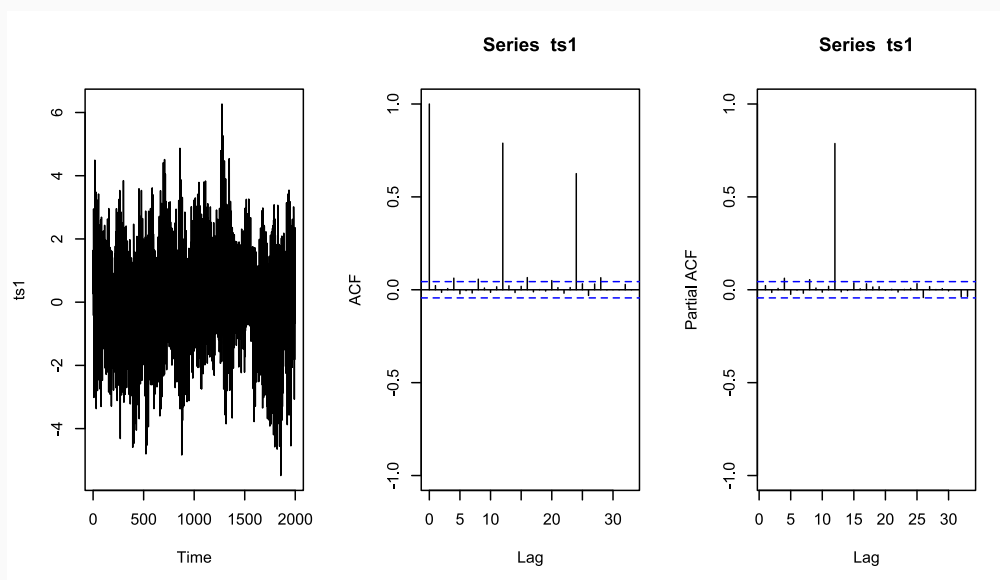
Seasonal ARIMA models(cont'd)

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF.
- **Seasonal ARIMA (0,0,0) (0,0,1)₁₂** will show:
 - a spike at lag 12 in the **ACF** but no other significant spikes.
 - The **PACF** will show exponential decay in the seasonal lags (lags 12, 24, ...).
 - Note: **ACF** starts at lag 0 and **PACF** starts at lag 1.



Seasonal ARIMA models(cont'd)

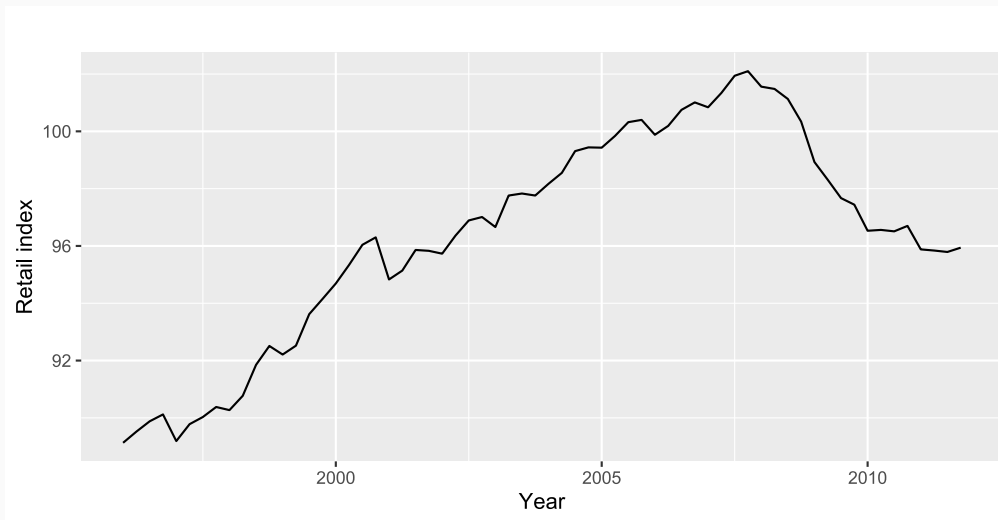
- **Seasonal ARIMA (0,0,0)(1,0,0)**₁₂ will show:
 - exponential decay in the seasonal lags of the **ACF**.
 - a single significant spike at lag 12 in the **PACF**.
- Note: **ACF** starts at lag 0 and **PACF** starts at lag 1



Data Example: European quarterly retail

- Quarterly retail trade index in the Euro area (17 countries), 1996-2011.

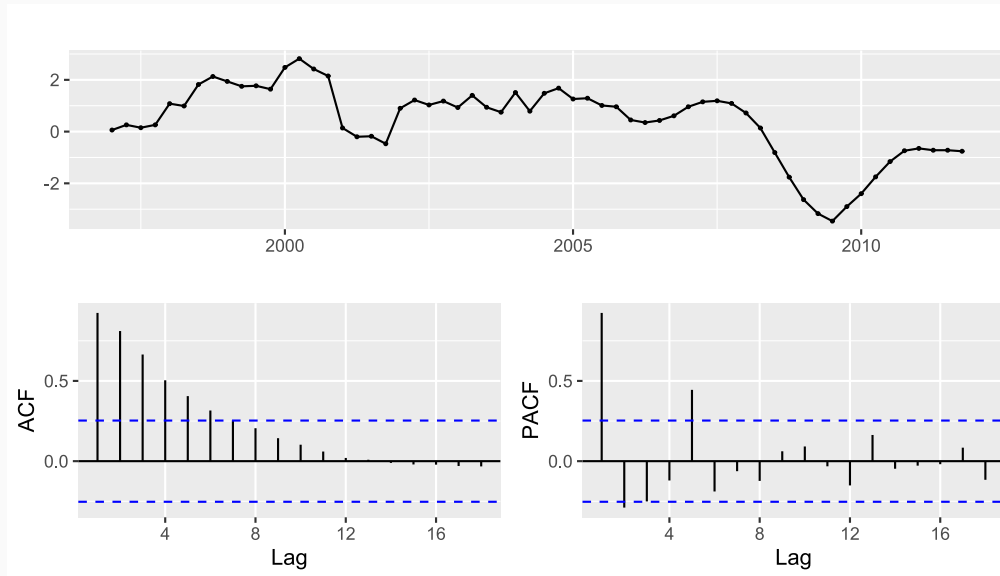
```
library(fpp2)
autoplot(euretail) +
  xlab("Year") + ylab("Retail index")
```



Data Example (cont'd)

- Taking the seasonal difference

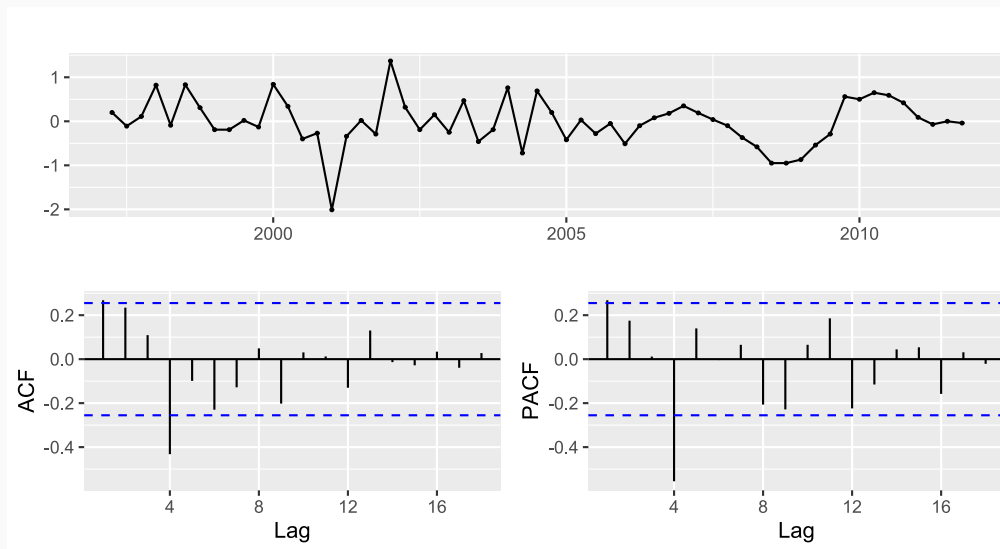
```
euretail %>% diff(lag=4) %>% ggtsdisplay()
```



Data Example (cont'd)

- Both seasonal and first differences are applied:

```
euretail %>% diff(lag=4) %>% diff() %>%  
  ggtsdisplay()
```



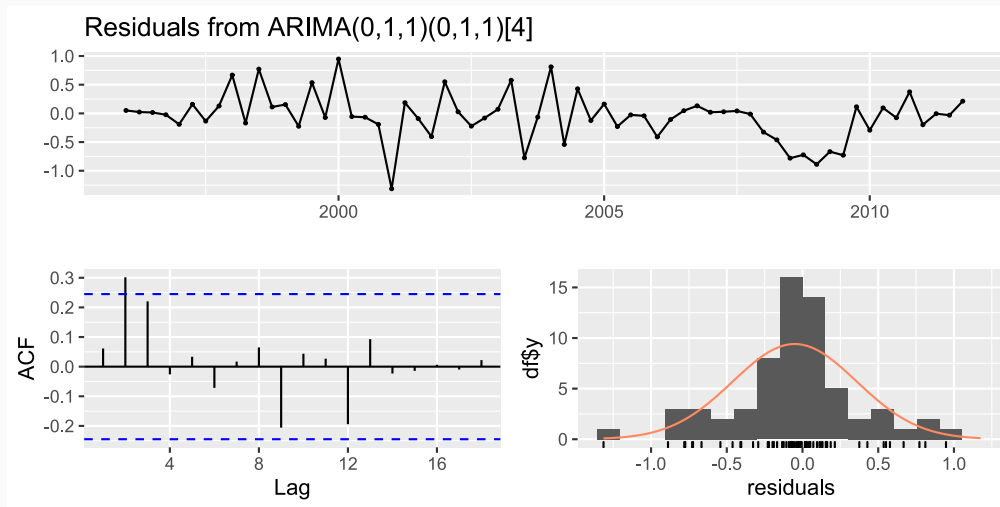
Data Example (cont'd)

- $d = 1$ and $D = 1$ seems necessary.
- Significant spike at lag 1 in ACF suggests non-seasonal MA(1) component.
- Significant spike at lag 4 in ACF suggests seasonal MA(1) component.
- Initial candidate model: $\text{ARIMA}(0,1,1)(0,1,1)_4$.
- We could also have started with $\text{ARIMA}(1,1,0)(1,1,0)_4$.

Data Example (cont'd)

- $\text{ARIMA}(0,1,1)(0,1,1)_4$

```
fit <- Arima(euretail, order=c(0,1,1),  
  seasonal=c(0,1,1))  
checkresiduals(fit)
```



```
##
```

```
##      Ljung-Box test
```

```
##
```

```
## data:  Residuals from ARIMA(0,1,1)(0,1,1)[4]
```

```
## Q* = 10.654, df = 6, p-value = 0.09968
```

```
##
```

Data Example (cont'd)

- ACF and PACF of residuals show significant spikes at **lag 2**, and maybe **lag 3**.
- BIC of ARIMA(0,1, **3**)(0,1,1)₄ model is 77.64592.

```
fit <- Arima(euretail, order=c(0,1,3),  
  seasonal=c(0,1,1))  
BIC(fit)
```

```
## [1] 77.64592
```


Data Example (cont'd)

- BIC of ARIMA(0,1,2)(0,1,1)₄ model is 81.83933.

```
(fit2 <- Arima(euretail, order=c(0,1,2),  
  seasonal=c(0,1,1)))
```

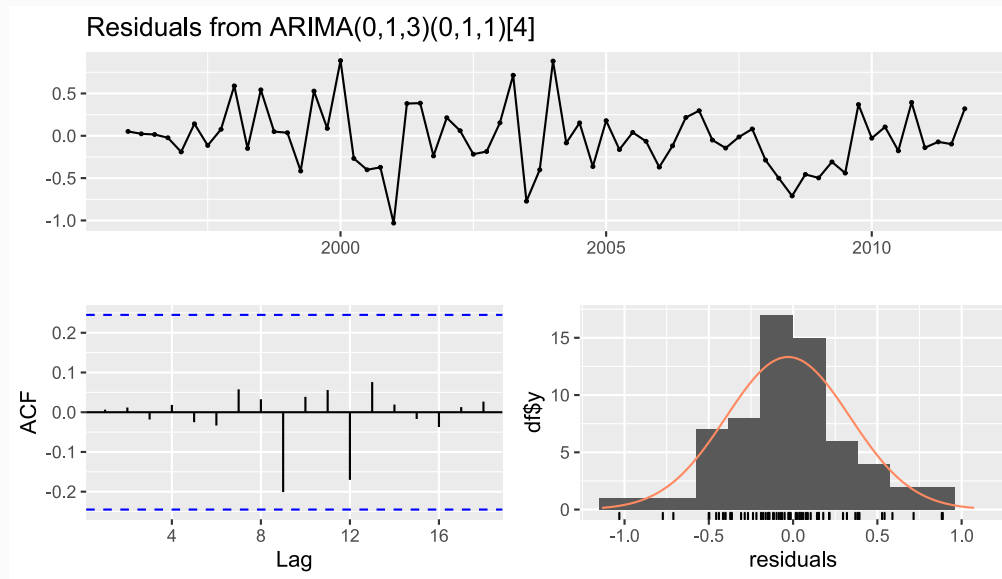
```
## Series: euretail  
## ARIMA(0,1,2)(0,1,1)[4]  
##  
## Coefficients:  
##          ma1      ma2      sma1  
##      0.2303  0.2502  -0.6991  
## s.e.  0.1484  0.1188  0.1284  
##  
## sigma^2 estimated as 0.1789:  log likelihood=-32.76  
## AIC=73.53   AICc=74.27   BIC=81.84
```

```
BIC(fit2)
```

```
## [1] 81.83933
```

Data Example (cont'd)

```
checkresiduals(fit)
```



```
##
```

```
##      Ljung-Box test
```

```
##
```

```
## data:  Residuals from ARIMA(0,1,3)(0,1,1)[4]
```

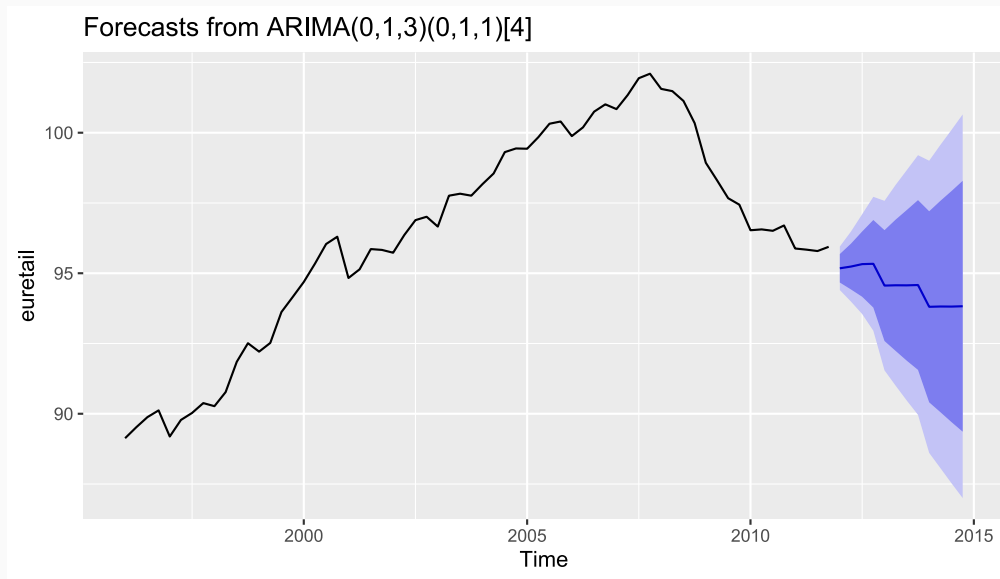
```
## Q* = 0.51128, df = 4, p-value = 0.9724
```

```
##
```

```
## Model df: 4.    Total lags used: 8
```

Data Example (cont'd)

```
autoplot(forecast(fit, h=12))
```



Data Example (cont'd)

```
auto.arima(euretail)
```

```
## Series: euretail
```

```
## ARIMA(0,1,3)(0,1,1)[4]
```

```
##
```

```
## Coefficients:
```

```
##          ma1      ma2      ma3      sma1
```

```
##          0.2630  0.3694  0.4200  -0.6636
```

```
## s.e.    0.1237  0.1255  0.1294   0.1545
```

```
##
```

```
## sigma^2 estimated as 0.156:  log likelihood=-28.63
```

```
## AIC=67.26   AICc=68.39   BIC=77.65
```

Data Example (cont'd)

```
auto.arima(euretail, stepwise=FALSE, approximation=FALSE)
```

```
## Series: euretail
## ARIMA(0,1,3)(0,1,1)[4]
##
## Coefficients:
##          ma1      ma2      ma3      sma1
##      0.2630  0.3694  0.4200  -0.6636
## s.e.  0.1237  0.1255  0.1294   0.1545
##
## sigma^2 estimated as 0.156:  log likelihood=-28.63
## AIC=67.26   AICc=68.39   BIC=77.65
```